



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**25.09.2002 Bulletin 2002/39**

(51) Int Cl.<sup>7</sup>: **C07K 14/71, C12N 15/12,  
G01N 33/53**

(21) Application number: **02006616.3**

(22) Date of filing: **25.03.2002**

(84) Designated Contracting States:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE TR**  
Designated Extension States:  
**AL LT LV MK RO SI**

- **Parge, Hans**  
**San Diego, CA 92103 (US)**
- **Sarup, Jay**  
**San Diego, CA 92122 (US)**
- **Zhu, Jeff**  
**Carlsbad, CA 92009 (US)**

(30) Priority: **23.03.2001 US 277968 P**

(71) Applicant: **Agouron Pharmaceuticals, Inc.**  
**La Jolla, CA 92037 (US)**

(74) Representative: **Keller, Günter, Dr. et al**  
**Lederer & Keller**  
**Patentanwälte**  
**Prinzregentenstrasse 16**  
**80538 München (DE)**

(72) Inventors:  
• **Mroczkowski, Barbara**  
**Encinitas, CA 92024 (US)**  
• **Hickey, Michael**  
**San Diego, CA 92126 (US)**  
• **McTigue, Michele Ann**  
**Encinitas, CA 92024 (US)**  
• **Murray, Brion William**  
**San Diego, CA 92106 (US)**

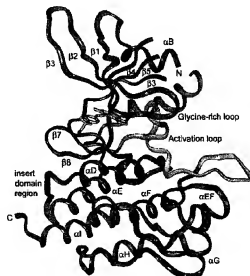
Remarks:

Das Sequenzprotokoll, das als Anlage zu den Anmeldungsunterlagen mitveröffentlicht ist, ist nach dem Anmeldetag eingereicht worden. Der Anmelder hat erklärt, dass dieses nicht über den Inhalt der Anmeldung in der ursprünglich eingereichten Fassung hinausgeht.

(54) **Catalytic domains of the human hepatocyte growth factor receptor tyrosine kinase and methods for identification of inhibitors thereof**

(57) The identification, isolation, purification, and characterization of the catalytic domain of the human hepatocyte growth factor receptor kinase (hHGFR) are described. A crystal structure of the hHGFR kinase domain is reported herein. This structure provides a three-dimensional description of the binding site of the hHGFR for structure-based design of small molecule inhibitors thereof as therapeutic agents. The kinase domain of human HGFR and its associated crystal structure is described for use in the discovery, identification and characterization of modulators of human HGFR.

**Figure 2**



## Description

## CROSS-REFERENCES TO RELATED APPLICATIONS

- 5 **[0001]** This application claims the benefit of U.S. provisional application Serial No. 60/277,968, filed March 23, 2001, which is hereby incorporated herein by reference in its entirety.

## FIELD OF THE INVENTION

- 10 **[0002]** The present invention generally relates to the isolation and purification of the catalytic domain of the human hepatocyte growth factor receptor kinase (the *Met* protooncogene product, *Met*; HGFR) and its use in the discovery, identification and characterization of inhibitors of same. The present invention further relates to the field of crystallography and, particularly, to X-ray crystallography data useful for identification and construction of therapeutic compounds in the treatment of various disease conditions associated with the *Met* receptor tyrosine kinase. More specifically, the  
15 invention relates to crystallized complexes of *Met*.

## BACKGROUND OF THE INVENTION

- [0003]** Hepatocyte growth factor (HGF), also known as scatter factor, is a mesenchymally derived cytokine capable of inducing a variety of pleiotropic effects in normal and neoplastic cells (Sonnenberg et al., *J. Cell Biol.* 123:223-235 (1993); Matsumoto et al., *Crit. Rev. Oncog.* 3:27-54 (1992); and Stoker et al., *Nature* 327:239-242 (1987)). These include proliferation of different types of epithelial and endothelial cells, dissociation of epithelial colonies into individual cells, stimulation of the motility (scattering) of epithelial cells, induction of epithelial morphogenesis (Montesano et al., *Cell* 67:901-908 (1991)), angiogenesis (Bussolino et al., *J. Cell Biol.* 119:629-641 (1992)), and promotion of the invasion of extracellular matrices (Stella et al., *Int. J. Biochem. Cell Biol.* 12:1357-62 (1999) and Stuart et al., *Int. J. Exp. Path.* 81: 17-30 (2000)). *In vivo*, HGF is involved in tissue regeneration, tumor invasion, and embryonic processes, all of which are dependent on both cell motility and proliferation.

- [0004]** HGF initiates these physiologic processes through a high affinity receptor identified as the *c-Met* protooncogene product (Park et al., *Proc. Natl. Acad. Sci. USA* 84:6379-83 (1987); and Boltaro et al., *Science*, 251:802-4 (1991)). The mature form of the receptor (HGFR) consists of an extracellular  $\alpha$ -subunit and a transmembrane  $\beta$ -subunit containing intrinsic tyrosine kinase activity. Engagement of the receptor induces dimerization which in turn up-regulates kinase activity. Activation of *Met* promotes transphosphorylation of several key tyrosine residues responsible for initiating downstream signaling cascades by recruiting multiple effectors (Furge et al., *Oncogene* 19:5582-9 (2000)). These include the p85 subunit of PI3-kinase, phospholipase C $\gamma$  (Gaul et al., *Oncogene* 19:1509-18 (2000)), Grb2 and Shc adaptor proteins, the protein phosphatase SHP2 and Gab1. The latter adapter has emerged as the major downstream docking molecule that becomes tyrosine phosphorylated in response to ligand occupancy (Schaeper et al., *J. Cell Biol.* 149:1419-32 (2000); Bardelli, et al., *Oncogene* 18:1139-46 (1999); and Sachs et al., *J. Cell Biol.* 150:1375-84 (2000)). Activation of other signaling molecules has been reported in HGF stimulated cells, most notably, Ras, MAP kinases and FAK (Tanimura et al., *Oncogene* 17:57-65 (1998); and Lai et al., *J. Biol. Chem.* 275:7474-80 (2000)). The role for many of these signaling molecules has been established in cell proliferation but is not as evident in cell dissociation and scattering.

- [0005]** The hepatocyte growth factor receptor (HGFR) is expressed predominantly in epithelial cells but has also been detected in endothelial cells, myoblasts, hematopoietic cells and motor neurons. Inappropriate activation of the receptor is implicated in the onset and progression of a number of tumors and in the promotion of metastasis. A direct link between HGFR and cancer has been shown by the identification of missense mutations in the Kinase domain which predispose individuals to papillary renal carcinomas (PRC) and hepatocellular carcinomas (HCC) (Giordano et al., *FASEB J* 14:399-406 (2000)).

- [0006]** Activation of this tyrosine kinase plays a key role in the regulation of migration, invasion and angiogenesis in cancer. The receptor is overexpressed in a significant percentage of human cancers and is amplified during the transition between primary tumors and metastasis. Missense mutations in the tyrosine kinase domain of the gene have been reported in the germline of affected members of PRC and HCC families (Park et al., *Cancer Res.* 59:307-10 (1999); Schmidt et al., *Nature Genetics* 16:68-73 (1997); and Schmidt et al., *Cancer Research* 58:1719-22 (1998)). Most of these genetic lesions represent disease-producing mutations that appear to accelerate carcinogenesis by constitutively activating the receptor. In addition, *in vivo* experiments indicate that autocrine HGF-*Met* signaling plays a significant role in the development and progression of certain malignancies (Belluscio et al., *Oncogene* 9:1091-99 (1994) and Rong et al., *Proc. Natl. Acad. Sci. USA* 91:4731-4735 (1994)). It is becoming increasingly evident that the *Met* signaling pathway is involved in the invasive behavior of various tumors by promoting not only tumor spreading, but also neovascularization (Ramirez et al., *Clin. Endocrinol.* 53:635-44 (2000)). Thus, selective, small molecule kinase

modulators are expected to have therapeutic potential for the treatment of cancers in which *Met* receptor activation plays a critical role in the development and progression of primary tumors and secondary metastases. Since HGF is also a known angiogenic, there is the potential for this class of modulators to impact angiogenesis-dependent diseases such as diabetic retinopathy.

[0007] A direct role for HGFR in the metastatic behavior of human malignancy has been documented in the literature since its initial identification as the cellular homologue of the *tpv-Met* oncogene (Cooper et al., *Nature* 311:29-33 (1984)). The receptor is overexpressed in various tumors including thyroid, ovarian and pancreatic carcinomas and is amplified in liver metastases of colorectal carcinomas (Rong et al., *Cancer Res.* 55:1963-1970 (1995); Rong et al., *Cancer Res.* 53:5355-5360 (1993); Kenworthy et al., *Br. J. Cancer* 66:243-247 (1992); and Scarpino et al., *J. Pathology* 189:570-575 (1999)). In patients with invasive breast carcinoma, expression of either the receptor or ligand is a predictor of decreased survival, further linking *Met* to tumor progression (Camp et al., *Cancer* 86:2259-65 (1999)). In general, most human tumors and tumor cell lines of mesenchymal origin inappropriately express HGFR and/or HGF. This observation supports the premise that HGFR plays a key role in sarcomagenesis and that both autocrine and paracrine signaling modes contribute to the development of human tumors of mesenchymal origin.

[0008] HGFR was originally identified as the cellular counterpart of the *tpv-Met* oncogene, the product of a chromosomal rearrangement generating a chimeric gene by fusing a leucine zipper motif to the tyrosine kinase domain of HGFR (Cooper et al., *Nature* 311:29-33 (1984)). The *tpv-Met* oncogene is activated via the leucine zipper interaction that is responsible for deregulating the enzymatic activity of the kinase domain. Accordingly, the *tpv-Met* oncogene efficiently transforms NIH-3T3 fibroblasts and transgenic expression of this oncogene leads to the development of hyperplasia and tumors in mice (Liang et al., *J. Clin. Invest.* 97:2872-2877 (1996)).

[0009] Almost a decade ago, Vande Woude's lab made the observation that mouse NIH 3T3 fibroblasts that overexpress *Met* can induce tumor formation in nude mice via an autocrine mechanism resulting in the interaction between recombinant *Met* receptor and endogenously expressed ligand (Rong et al., *Proc Natl Acad Sci USA* 91:4731-4735 (1994)). They also showed the transformed cells to be metastatic. Since then numerous reports have substantiated the initial observation that chronic *Met*-HGF signaling can induce tumor formation (Oda et al., *Human Pathology* 31:185-192 (2000)). For example, spontaneously transformed tumor cells, which express both ligand and receptor, routinely exhibit increased proliferation and motility. The co-expression of HGF and HGFR is common among non-small-cell lung cancers, especially adenocarcinoma. Not surprisingly, retroviral transduction of HGF in NCI-H358 lung adenocarcinoma cells that express HGFR endows these cells with enhanced capacity to colonize soft agar medium and to form xenograft tumors when implanted in immune-deficient mice (Seung et al., *Neoplasia* 2:228-234 (2000)).

[0010] The most compelling data linking HGFR signaling to human malignancies is the discovery of germline missense mutations that map to the kinase domain of HGFR in the majority of hereditary papillary renal cell carcinomas and the detection of somatic missense HGFR mutations in sporadic papillary kidney carcinomas and childhood hepatocellular carcinomas. These mutations which render the receptor constitutively active have been shown to confer an invasive phenotype to transfected cells (Jeffers et al., *Proc. Natl. Acad. Sci. USA* 94:11445-11450 (1997)).

[0011] The introduction of these mutations into wild-type *Met* cDNA results in transforming, tumorigenic, and metastatic properties in mouse cell lines. When these same mutations are introduced into mice as transgenes, the founders develop tumors that metastasize to secondary sites. Furthermore, it has been observed that cells carrying these *Met* mutations undergo clonal expansion during HNSCC (head and neck squamous cell carcinoma) progression, further correlating *Met* with the progression of primary cancers to metastasis (Renzo et al., *Oncogene* 19:1547-1555 (2000)).

[0012] Direct experimental evidence linking the *Met*-HGF signaling pathway to tumor cell metastasis has been validated in the mouse, using either transfected cells or transgenic animals (Takayama et al., *Proc. Natl. Acad. Sci. USA* 94:701-706 (1997)). For example, in rigorously controlled experiments, anti-*Met* oligonucleotides inhibit the proliferation and invasiveness of human gastric cancer cells (Kaji et al., *Cancer Gene Therapy* 3:393-404 (1998)). Dominant negative *Met* has been shown by numerous laboratories to reduce tumorigenicity and spontaneous metastasis both *in vitro* and *in vivo* (Firon et al., *Oncogene* 19:2386-2397 (2000)). *In vivo* a dramatic reduction in tumor and metastasis formation, accompanied by improved survival, was noted. Furthermore, peptides corresponding to the multifunctional docking site at the carboxy terminal tail of the *Met* receptor, that bind the receptor and inhibit kinase activity, inhibit HGF-mediated invasive growth, as measured by cell migration, invasiveness, and branched morphogenesis (Bardelli et al., *J. Biol. Chem.* 274:29274-29281 (1999)).

[0013] Naturally occurring splice variants of HGF have also been identified that behave as competitive antagonists of mature HGF (Date et al., *Oncogene* 17:3045-3054 (1998)). These variants inhibit autophosphorylation of the receptor and consequently block HGF-induced migration of human umbilical vein endothelial cells in a migration assay and in an endothelial wounding assay (Jiang et al., *Clinical Cancer Research* 5:3695-3703 (1999)). HGF antagonists have also been reported to inhibit the motility and invasion of colon, gallbladder and cervical carcinoma cells. Most significantly, infusion of these antagonists into nude mice implanted with tumor cells represses the invasion of tumorigenic cells into surrounding tissues (Kuba et al., *Cancer Research* 60:6737-6743 (2000)). Similarly, a monoclonal antibody highly selective for HGF has also been found to block tumor progression *in vitro* and *in vivo*. Cao et al., *Proc.*

*Natl. Acad. Sci. USA* 98:7443-8 (2001).

[0014] Recently, the geldanamycin family of anisamycin antibiotics has been implicated in the down-regulation of the HGFR at nanomolar concentrations. The loss of HGFR expression observed at nanomolar concentrations not only inhibits HGF-induced cell motility and invasion but also reverts the Met-transformed phenotype (Webb et al., *Cancer Research* 60: 342-349 (2000)). This class of compounds are currently in clinical trials (NCI) as potential anti-invasive, anti-metastatic agents.

[0015] PCT International Publication No. WO 01/09159 discloses nucleic acid ligands to HGF and HGFR. These ligands were isolated using SELEX (Systematic Evolution of Ligands Exponential enrichment). U.S. Patent Nos. 5,686,292 and 6,099,841 report HGFR antagonists and agonists, respectively. U.S. Patent No. 6,174,889 discloses bicyclic heteroaromatic compounds as protein tyrosine kinase inhibitors. PCT International Publication Nos. WO 00/43373, WO 98/07695 and WO 99/15550 disclose protein kinase inhibitor compounds.

[0016] Several attempts have been made to elucidate three dimensional structures of proteins to design candidate drugs (See, e.g., Davis et al., *Science* 291:134-137 (2001); Zhu et al., *Structure* 7: 651-661 (1999); and Ymaguchi et al., *Nature* 384: 484-489 (1996)). Further, PCT International Publication No. WO 00/70030 discloses the three-dimensional crystal structure of Lck with its ligand.

#### SUMMARY OF THE INVENTION

[0017] The generation, kinetic characterization, and structure determination of the kinase domain of the human HGFR protein is disclosed herein. The domain begins between residues 1051 and 1078 and terminates between residues 1341 and 1348 of the full-length protein [SEQ ID NO: 2]. The domain preferably extends from residues 1051-1341, and more preferably from residues 1051-1348. In one embodiment of the present invention, the domain has the sequence selected from the group of SEQ ID NOS: 3 through 5, 9, 13, 15, and 16.

[0018] In one of its aspects, the present invention relates to an isolated polynucleotide that encodes the human hepatocyte growth factor receptor or the human hepatocyte growth factor receptor kinase domain, or a fragment or variant thereof. In one embodiment, the nucleotide sequence of the polynucleotide corresponds to at least bases 3342 to 4206 of SEQ ID NO: 1. In other embodiments, the nucleotide sequence of the polynucleotide corresponds to the sequence of SEQ ID NOS: 10, 11, 12, or 14.

[0019] In another of its aspects, the present invention relates to a crystal structure containing the human hepatocyte growth factor receptor kinase. In one embodiment, the amino acid sequence of the kinase corresponds to at least amino acids 1051 to 1348 of SEQ ID NO: 2. In other embodiments, the amino acid sequence of the kinase corresponds to the sequence of SEQ ID NOS: 3, 4, 5, 6, 7, 8, 9, 13, 15, or 16.

[0020] In still another of its aspects, the present invention relates to an isolated polypeptide containing the human hepatocyte growth factor receptor or human hepatocyte growth factor receptor kinase domain, or a variant thereof. In one embodiment, the human hepatocyte growth factor receptor or human hepatocyte growth factor receptor kinase domain contains a deletion that imparts favorable physical characteristics to the resulting polypeptide (e.g., suitability for analysis by nuclear magnetic resonance, suitability for high throughput screening, suitability for biochemical characterizations, suitability for x-ray crystallography, suitability for colorimetry and suitability for other diagnostic methods). In other embodiments, the polypeptide contains amino acids 1051 to 1341 of the sequence as set forth in SEQ ID NO: 2; amino acids 1051 to 1348 of the sequence as set forth in SEQ ID NO: 2; or the amino acid sequence as set forth in SEQ ID NOS: 3, 4, 5, 6, 7, 8, 9, 13, 15, or 16; or a conservatively substituted variant thereof.

[0021] In yet another of its aspects, the present invention relates to an isolated polynucleotide that encodes the catalytically active form of the human hepatocyte growth factor receptor or human hepatocyte growth factor receptor kinase domain, or a fragment or variant thereof.

[0022] A further aspect of the present invention relates to an isolated catalytically active polypeptide comprising the human hepatocyte growth factor receptor or human hepatocyte growth factor receptor kinase domain, or a variant thereof.

[0023] Another aspect of the present invention relates to an isolated polynucleotide which encodes the catalytic domain of the human hepatocyte growth factor receptor kinase, or a fragment or variant thereof.

[0024] Still another aspect of the present invention relates to an isolated catalytically active polypeptide containing the catalytic domain of the human hepatocyte growth factor receptor kinase or a variant thereof.

[0025] Yet another aspect of the present invention relates to an isolated soluble polypeptide comprising the catalytic domain of the human hepatocyte growth factor receptor kinase or a variant thereof.

[0026] The present invention also relates to an expression vector for producing the human hepatocyte growth factor receptor kinase in a host cell. The vector contains a polynucleotide encoding the human hepatocyte growth factor receptor kinase or a variant thereof; and regulatory sequences that are functional in the host cell and operably linked to the polynucleotide. In one embodiment, the polynucleotide encodes the active human hepatocyte growth factor receptor kinase containing bases 3342 to 4206 of SEQ ID NO: 1. In another embodiment, the vector is selected from

the group consisting of pET28a, pAcSG2, and pFastBac. In a further embodiment, the host cell is *E. coli*.

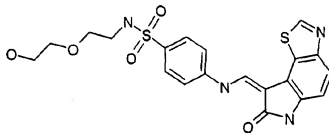
[0027] In another aspect, the present invention relates to a host cell transformed or transfected with a polynucleotide encoding the human hepatocyte growth factor receptor kinase or a variant thereof. In one embodiment, the host cell is transformed or transfected with the polynucleotide via an expression vector containing the polynucleotide; a regulatory sequence functional in the host cell operably linked to the polynucleotide; and a selectable marker. The expression vector can be, for example, pET28a, pAcSG2, and pFastBac. In another embodiment, the polynucleotide encodes the human hepatocyte growth factor receptor kinase containing bases 3342 to 4206 of SEQ ID NO: 1. In a further embodiment, the host cell is *E. coli*. In yet another embodiment, the host cell is infected with a recombinant baculovirus. Additionally, the host cell can be insect cell, such as Sf9.

[0028] The present invention additionally relates to a method of producing a polypeptide or variant thereof by culturing a host cell, transformed or transfected with a polynucleotide encoding the human hepatocyte growth factor receptor kinase or a variant thereof, under conditions such that the polypeptide or variant thereof is expressed; and recovering the polypeptide or variant.

[0029] In yet another of its aspects, the present invention relates to a method for assaying a candidate compound for its ability to interact with the human hepatocyte growth factor receptor. The method involves expressing an isolated DNA sequence or variant thereof encoding the kinase domain of the human hepatocyte growth factor receptor in a host capable of producing the kinase in a form which may be assayed for interaction with the candidate compound. The kinase is exposed to the candidate compound and the interaction of the kinase with the candidate compound is evaluated. In one embodiment, the interaction is evaluated by crystallizing the kinase in a condition suitable for x-ray crystallography; and conducting x-ray crystallography on the kinase. The results of the x-ray crystallography are optionally used to determine the three dimensional molecular structure of the configuration of human hepatocyte growth factor receptor kinase and the binding pockets thereof.

[0030] The present invention further relates to a crystal structure containing a polypeptide encoded by a polynucleotide which encodes the human hepatocyte growth factor receptor kinase domain, or a fragment or variant thereof.

[0031] In addition, the present invention relates to a crystal structure containing a polypeptide encoded by a polynucleotide which encodes the human hepatocyte growth factor receptor kinase domain, or a fragment or variant thereof, and a ligand complexed thereto. In one embodiment, the ligand modulates the activity of human hepatocyte growth factor kinase. In another embodiment, the ligand is a compound of the formula:



[0032] The present invention still further relates to a process of drug design for compounds which interact with the human hepatocyte growth factor receptor kinase. The process involves crystallizing the human hepatocyte growth factor receptor kinase and resolving the x-ray crystallography of the kinase. The data generated from resolving the x-ray crystallography of the kinase is then applied to a computer algorithm which generates a model of the kinase suitable for use in designing molecules that will act as agonists or antagonists to the polypeptide. An iterative process is then applied whereby various molecular structures are applied to the computer-generated model to identify potential agonists or antagonists of the kinase. In one embodiment, the process is utilized to identify modulators of the active kinase, which serve as lead compounds for the design of potentially therapeutic compounds for the treatment of diseases or disorders associated with the hepatocyte growth factor receptor- hepatocyte growth factor signaling pathway.

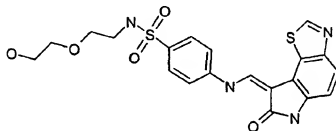
[0033] In yet another aspect, the present invention relates to a method of rapidly screening large compound libraries to identify compounds that inhibit human hepatocyte growth factor receptor kinase containing a non-radioactive immunosorbent assay capable of robotic control. In one embodiment, the assay is DELFIA.

[0034] In still another aspect, the present invention relates to a method of assessing compounds which are agonists or antagonists of the activity of the hepatocyte growth factor receptor kinase by crystallizing the hepatocyte growth factor receptor kinase and obtaining crystallography coordinates for the crystallized hepatocyte growth factor receptor

kinase. The crystallography coordinates are then applied to a computer algorithm such that the algorithm generates a model of the kinase suitable for use in designing molecules that will act as agonists or antagonists to the kinase. An iterative process is used to apply various molecular structures to the computer-generated model to identify potential agonists or antagonists to the kinase. The agonist or antagonist is then optionally synthesized or obtained, and contacted with the molecule to determine the ability of the potential agonist or antagonist to interact with the molecule.

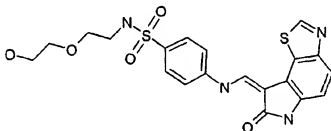
[0035] The present invention also relates to a method for determining the three-dimensional structure of a complex of hepatocyte growth factor receptor kinase with a ligand, wherein x-ray diffraction data for crystals of the complex are obtained, and the set of atomic coordinates of Table 1 or portions thereof, and coordinates having a root mean square deviation therefrom with respect to conserved protein backbone atoms of not more than about 1.5 Å are used to define the three-dimensional structure of the complex.

[0036] In a further of its aspects, the present invention relates to a method of using a three-dimensional structure of a polypeptide encoded by a polynucleotide which encodes the human hepatocyte growth factor receptor and a compound of the formula:



as defined by the structure coordinates of Table 1, or a portion thereof, in a drug-discovery strategy. A potential drug is selected, in conjunction with computer modeling, by performing rational drug design with the three-dimensional structure determined from one or more sets of atomic coordinates in Table 1. The potential drug is contacted with a polypeptide containing a functional human hepatocyte growth factor receptor and the binding of the potential drug with the polypeptide is determined.

[0037] The present invention still further relates to a method of using a three-dimensional structure of a polypeptide encoded by a polynucleotide which encodes the human hepatocyte growth factor receptor kinase domain and a compound of the formula:



as defined by the structure coordinates of Table 1, or a portion thereof, in a drug-discovery strategy. A potential drug is selected, in conjunction with computer modeling, by performing rational drug design with the three-dimensional structure determined from one or more sets of atomic coordinates in Table 1. The potential drug is contacted with a polypeptide containing a functional human hepatocyte growth factor receptor. Whether or not the potential drug modulates the activity of the polypeptide is then determined.

[0038] The present invention further relates to a method for evaluating the potential of a chemical entity to associate with either: (a) a molecule or molecular complex having a binding pocket defined by structure coordinates of human hepatocyte growth factor receptor amino acids 1082-1086, 1091-1094, 1107-1110, 1140-1142, 1155-1175, 1208-1213, and 1219-1231, according to Table 1, or (b) a homologue of the molecule or molecular complex having a binding pocket

that has a root mean square deviation from the backbone atoms of the amino acids of not more than about 1.5 Å. The method involves employing computational means to perform a fitting operation between the chemical entity and a binding pocket defined by structure coordinates of hepatocyte growth factor receptor amino acids 1082-1086, 1091-1094, 1107-1110, 1140-1142, 1155-1175, 1208-1213, and 1219-1231 which are within about a root mean square deviation of not more than about 1.5 Å from the backbone atoms of said amino acids according to Table 1; and analyzing the results of the fitting operation to quantify the association between the chemical entity and the binding pocket.

[0039] In yet another of its aspects, the present invention relates to a computer for producing a three-dimensional representation of: (a) a molecule or molecular complex, wherein said molecule or molecular complex comprises a binding pocket defined by the structure coordinates of hepatocyte growth factor receptor kinase amino acids 1082-1086, 1091-1094, 1107-1110, 1140-1142, 1155-1175, 1208-1213, and 1219-1231, according to Table 1; or (b) a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than about 1.5 Å. The computer comprises includes a computer-readable data storage medium having a data storage material encoded with computer-readable data containing the structure coordinates of hepatocyte growth factor kinase amino acids 1082-1086, 1091-1094, 1107-1110, 1140-1142, 1155-1175, 1208-1213, and 1219-1231, according to Table 1. The computer also includes a working memory for storing instructions for processing the computer-readable data; a central-processing unit coupled to the working memory and to the computer-readable data storage medium for processing the computer-machine readable data into the three-dimensional representation; and a display coupled to the central-processing unit for displaying the three-dimensional representation. In one embodiment, the computer produces a three-dimensional representation of: (a) a molecule or molecular complex defined by structure coordinates of all of the hepatocyte growth factor kinase amino acids set forth in Table 1, or (b) a homologue of the molecule or molecular complex having a binding pocket that has a root mean square deviation from the backbone atoms of the amino acids of not more than 1.5 Å.

[0040] The present invention also relates to a computer for determining at least a portion of the structure coordinates corresponding to the x-ray diffraction data obtained from a molecule or molecular complex. The computer includes a computer-readable data storage medium having a data storage material encoded with machine-readable data, wherein the data includes at least a portion of the structural coordinates of hepatocyte growth factor receptor kinase according to Table 1. The computer also includes a computer-readable data storage medium having a data storage material encoded with computer-readable data including x-ray diffraction data obtained from the molecule or molecular complex; a working memory for storing instructions for processing the computer-readable data; a central-processing unit coupled to the working memory and to the computer-readable data storage medium for performing a Fourier transform of the machine readable data and for processing the computer-readable data into structure coordinates; and a display coupled to the central-processing unit for displaying the structure coordinates of the molecule or molecular complex.

[0041] In still another aspect, the present invention relates to a computer readable medium having stored thereon data of the structure coordinates of a *Mef* ligand-binding site including 1082-1086, 1091-1094, 1107-1110, 1140-1142, 1155-1175, 1208-1213, and 1219-1231 according to Table 1.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0042] The numerous objects and advantages of the present invention may be better understood by those skilled in the art by reference to the accompanying detailed description and the following drawings. The application file contains at least one drawing executed in color. Copies of this patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

Figure 1 is a purification scheme of the human HGFRkd;

Figure 2 is a ribbon representation of the kinase domain of the HGFR protein structure with Compound 1 bound thereto, wherein the N- and C-termini are indicated by N and C, respectively. Colors: Compound 1 (purple), Glycine-rich loop (orange), activation loop (yellow), alpha helix C (green), kinase insert domain (red), and remainder of protein (blue);

Figure 3 is a ribbon representation of the kinase domain of the HGFR kinase activation loop. Colors: activation loop (purple), Glycine-rich loop (yellow), alpha helix C (red), Phenylalanine 1089 (light blue), Aspartic acid 1228 (green), and Tyrosine-1230 (dark blue);

Figure 4 is an atomic representation of the Compound 1-HGFR kinase domain binding area, wherein the positions of bound water molecules are shown as red crosshairs. Colors: carbon (green), nitrogen (blue), oxygen (red), and sulfur (yellow);

Figure 5(A) is a Coomassie stained isoelectric focussing (IEF) electrophoretic evaluation of a time-course (20 °C) of HGFR autophosphorylation;

Figure 5(B) is an autoradiogram of IEF gel;

Figure 5(C) is a kinetic evaluation of the activation time course at 4°C;

Figure 5(D) is a MALDI-TOF evaluation of HGFR and pHGFR peptides derived from an exhaustive tryptic digest;

Figure 5(E) is a parent ion scan by nano-ESI-MS trypsin proteolyzed HGFR;

Figure 6(A) is a graph showing inhibition of HGFR and pHGFR by Compound 1 as measured in the coupled enzymatic assay; and

Figure 6(B) is a graph showing double reciprocal analysis of Compound 1 inhibition of pHGFR.

#### DETAILED DESCRIPTION OF THE INVENTION AND PREFERRED EMBODIMENTS

[0043] The terms "comprising" and "including" are used herein in their open, non-limiting sense.

[0044] The catalytic domain of the human HGFR receptor kinase facilitated crystallography, enzyme characterization, and high throughput screening of inhibitors. In particular, the catalytic domain of the HGFR kinase domain was used to determine its three-dimensional structure, which provides unique structural information useful for drug design.

[0045] As used herein, the abbreviation "HGFR" or "Met" refers to the polynucleotide encoding the hepatocyte growth factor receptor tyrosine kinase, or the protein *per se*. Also, the abbreviation "hHGFR" or "human Met", as used herein, refers to the polynucleotide encoding the human hepatocyte growth factor receptor tyrosine kinase or the protein *per se*. The HGFR protein is sometimes referred to as HGFR tyrosine kinase or HGFR kinase throughout the application. The nucleic acid sequence of the polynucleotide encoding the full-length protein of hHGFR was published by Park et al. (Proc. Natl. Acad. Sci. USA 84: 6379-83 (1987)) and submitted to GenBank under the accession number NM\_000245. The nucleic acid sequence described therein is provided herein, as SEQ ID NO: 1. The corresponding peptide sequence of the full-length protein is provided herein, as SEQ ID NO: 2. This peptide sequence was submitted to GenBank by Park et al. and assigned Accession number P08581. The intracellular domain of HGFR is provided herein as SEQ ID NO: 12.

#### SEQ ID NO. 1:

```
cgccctcgcc gccgcggcgc ccccgagcgc ttgtgagca gatcggagc cagtgaggg
gcgcgagcca gatcggggc gacagctgac ttgtgagag gagcgggga ggcgggagc
gcgcgtggg tcttgcgcc gctgacttc cactgtgtc ctgggcaccg aagatataac
ctctcataat gaaggccccc gctgtgttg cactggcat cctcgtgctc ctgtttacct
tggtgcagag tagcaatggg gagtgtaaag aggcactagc aaagtccgag atgaatgtga
ataagagta tcagctccc aacttcaccg cggaaacacc caccagaat gtcatttac
atgagatca catittcctt ggtgccacta actacattta tgtttaaat gaggaagacc
ttcagaagg ttctgagtac aagactggc ctgtgtgtga acaccagat tgtttccat
gtcaggactg cagcagcaaa gccaatatt caggaggtgt ttggaagat aatcaaca
tgctctagt tctgcacac tactatgatg atcaactat tagctgtgc agcgtcaaca
gagggacctg ccagcgacat gctttcccc acaatcatac tgcgtgata cagtcggagg
ttcactgcat atttcccca cagatagaag agcccagcca gtgcctgac tgtgtgtga
gcgccttggg agcctaagtc ctttcatcg taaaggaccg gttcatcaac tctttgtag
gcaalaccat aaattctct tatttccag atcatccatt gattcgata tcagtgaga
ggctaaagga aacgaagat ggtttatgt tttagcgga ccagtcctac attgatgitt
tacctgagtt cagagattct taccacatta agtatgccat tgctttgaa agcaacaatt
```



ttatttactt ctgacggc caaagggaaa ctctagatgc tcagactttt cacacaagaa  
 taacaggtt ctgtccata aactcggat tgcattccta catggaaatg cctctggagt  
 gtattctcac agaaaaagaa aaaaagagat ccacaaagaa ggaagtgtt aatatacttc  
 5 aggcctcgta tgcagcaag cctggggccc agcttctgag acaaatagga gccagcctga  
 atgatgacat tcttttggg gtgttcgcac aaagcaagcc agattctgcc gaaccaatgg  
 atcgalctgc catgttgca ttccctatca aatagtcaaa cgacttcttc aacaagatgc  
 tcaacaaaaa caatgtgaga tgcctccagc attttacgg acccaatcat gagcactgct  
 10 ttaataggac actcttgaga aattcatcag cctgtgaagc gcgccgtgat gaatacga  
 cagagtttac cacagcttg cagcgcgttg acttattcat gggtaacttc agcgaagtcc  
 tcttaacatc tatalccacc ttcatlaaag gagacctcac catagctaat ctggggacat  
 cagagggtcg ctctatgcag gtgtgtgttt ctgcatcagg accataaacc cctcatgtga  
 atttttccct ggactcccat ccagtgtctc cagaagtgat tgtggagcat acattaaacc  
 15 aaaaatggcta cacactgggt atcactggga agaagatcac gaagatccca ttgaatggct  
 tggcgtcgag acatttccag tctcgcagtc aatgcctctc tgcctcacc ttgttcagt  
 gtggctgggt ccacgacaaa tgtgtgcgat cggggaagtg cctgagcggg acatggactc  
 aacagatctg tctgcctgca atctacaagg ttctccaaa tagtgcacc cttgaaggag  
 20 ggacaaggct gaccatagt ggcctgggact ttggatttc gaggaataat aatttgatt  
 taaagaaaac tagagtctc ctggaaaag agagctgcac ctgacttta agtgagagca  
 cgtatgaatc atgaaatgc acagtgttc ctgcctgaa taagcatttc aatagtcca  
 taattatttc aaatggccac gggacaacac aatcacgtac atttctcat gggacctc  
 taataacaag tatttgcgc aaatcggctc ctatggctgg tggcacttta ctactttaa  
 25 ctggaaaata cctaacaagt gggaaattca gacacatttc aattgttga aaaacatga  
 cttaaaaaag tgtgtcaaac agttattctg aatgttatcc cccagcccaa accatticaa  
 ctgagtgttc tgttaaatg aaaaatgact tagccaaccg agagacaagc atcttcagtt  
 accgtgaaga tcccatgtc tatgaattc atccaacaa atcttttatt agtacttgg  
 30 ggaagaaccc tctcaacatt gtcagttttc tatttgcct tgcagtggtt gggagacaaa  
 taacaggtgt tgggaaaaac ctgaattcag ttatgttccc gagaatgttc ataatgtgc  
 atgaagcagg aaggaaactt acagtggcat gtcaacatgc ctctaattca gagataatct  
 gtgtaccac tcttccctg caacagctga atctgaact ccccttgaaa accaaagcct  
 35 ttctcatgtt agatgggac ctctccaat actttgaict catttatga cataactctg  
 tgtttaaagg ttgtgaaaag ccagtgtga tctcaatggg caatgaaaa gtactggaaa  
 taaagggaaa tgaatagac cctgaagcag taaagggtga agtgttaaaa gtgggaaata  
 agagctgtga gaatatcac ttacattcgt aagccgtttt atgcacggtc cccaatgacc  
 40 tgcgtaaat gaacagcag ctcaatatag agtggaaaga agcaatttct tcaaccgtcc  
 ttggaagaat aatagticaa ccagatcaga atttcacagg attgattct ggtgtgtct  
 caatatcaac agcactgtta ttactacttg ggttttctc gtggctgaaa aagagaagc  
 aaattaaaga tctgggcagt gaattatgt cttacgatgc aagatgacac actctcatt  
 45 tggataggct tgaagtgcc cgaagtgtaa gcccaactac agaaaatgtt tcaaatgaat  
 ctgtagacta ccgagctact ttccagaag atcagtttcc taattcatct cagaacggtt  
 catgccgaca agtgcagat cctctgacag acatgtcccc catcttaact agtggggact  
 ctgatalatc cagtcctatta ctgcaaaata ctgtccatc tgcactcagt gctctaaac  
 cagagctggt ccaggcagtg cagcatgtag tgaatggccc cagtgcctg atgtgcatt  
 50 tcaatgaagt cataggaaga gggcatlttg gtgtgtata tcatgggact ttgttgaca  
 atgatggcaa gaaaattcac tgtctgtga aatccttga cagaactact gacataggag  
 aagtttcca atttctgacc gagggaatca tcatgaaaga tttagtcat cccaatgac

5 tctcgtctct gggaaatctgc ctgcgaagtg aagggtctcc gctgggtgic ctaccatata  
 tgaacacatgg agatcttgcga aatttcattc gaaatgagac tcataatcca actgtaaaaag  
 aicctattgg ctttggctct caagtagcca aagcgaatgaa ataicttgca agcaaaaagt  
 ttgtccacag agacttggct gcaagaacct gtaictgga tgaanaatic acagicaagg  
 ttgctgattt tggcttggc agagacatgt atgataaaga atactatagt gtacacaaca  
 10 aaacaggtgc aaagctgcca gtgaagtggga tggcttggga aagctgcaa acicaaaaagt  
 ttaccaccaa gtacgatgtg tggctcttgg gcgtcgtcct ctgggagctg atgacaagag  
 gagccccacc ttatcctgac gtaaacacct ttgataaac tggttactg ttgcaaggga  
 gaagactcct acaacccgaa tactgcccag accccttata tgaagtaatg ctaaaatgct  
 ggcaccccaa agcgcgaatg cgcccatcct ttcttgaact ggtgtcccgc atatcagcga  
 tcttcttact ttatctggg gagcactatg tccatgtgaa cgctacttat gtgaacgtaa  
 15 aatgtgtgc tccgtatcct tctctgtgt catcagaaga taacgctgat galgagggtg  
 acacacgacc agcctccttc tgggagacat catagtecta gtactatgic aaagcaacag  
 tccacacttt gtccaatggt tttttactg cctgaccttt aaaaggccat cgatatctt  
 tgcctctgc cataggactt gtattgtat ttaaatctat ggattctaa gaaattctta  
 20 tctgacagag catcagaacc agaggcttgg tcccacaggc cagggaccaa tgcgtctcag

## SEQ ID NO. 2:

25 MKAPA VLAPGIL VLLFTLVQSRNGECKEALAKSEMNVNMKYQLPNFT AETPIQNVIL  
 HEHHIFLGATNYIYVLNEEDLQKV AEYKTGPVLEHPDCFPQDCSSKANLSGGVWVKD  
 NNMALV VDTY YDDQLISCGSVNRGT CQRHVFPNHNT ADIQSEVHCIFSQIEEPSQCP  
 DCVV SALGAKVLSS VKDRFINFFVGNNTNSSYFPDHLPLHSIS VRRLKETKD GFMFLTD  
 30 QSYID VLEFRDSYPIKY VHAFESNNFIYFLT VQRETLDAQIFHTRIIFCSINSGLHSY  
 MEMPLECILTEKKRKRSTKKEVFNLQAA YVSKPGAQLARQIGASLND DILFGVFAQS  
 KPDSAEPMDRSAMCAFPKIYVNDFFNKIVNKNVRCLQHFGPNHEHCFNRLLRNS  
 SGCEARRDEYRTEFTALQRVDLFMGQFSEVLLTSISTFIKGDLTIANLGTSEGRFMQV  
 VVSRSGPSTPHVNFLLDSHPVSPFIVEHTLNQNGYTLVTIGKKITKIFLNLGCRHFQ  
 35 SC SQCLSA PPFVQCGWCHDKCVRSEELSGTWQQICLPAIYKVFNPNSAPLEGGTRLT  
 ICGWDFGFRNRNKFLLKKTRVLLGNESCTLT LSESTMNTLKTCTVGPAMNKHFNMSIII  
 SNGHGTQYSTFSYVDPVITSISPKYGPMAGGTLTLTGNYLNSGNSRHISIGGKTC TL  
 KSVSNSILECYTPAQITISTEFAVKLKIDLANRETSIFS YREDPIVYEHPTKSFSISGGSTIT  
 40 GVGKNLNSVSVPRMVINVHEAGRNTVACQHRSNSEIICCTTPSLQQLNLQLPLKTKA  
 FFMLDGLSKYFDLIYVHNVPFKPFKEKPVMSMGNNENVLEIKGNIDIDPEAVKGEVLVK  
 GNKSCENHLHSEAVLCTVPNDLLKLNSELNIEWKQASISTVLGKVIVQPDQNFGLIA  
 GVSISTALLLLGFFLWLKRRKQIKDLGSELVRYDARVHTPHLDRLVSARSVSPTE  
 MVSNESVDYRATFPEDQFPNSSQNGSCRQVQYPLTDMSPILTS GDSIDISPLLQNTVHI  
 45 DLSALNPELVQAVQHV VIGPSSLIHVHNEVIGRGHFGCVYHGTLLDNDGKKIHCIAVK  
 SLNRITDIGEVSQFLTEGIIMKDFSHPNVLSLLGICLRSEGSPLVVLPMYMKHGDRLNFR  
 NETHNPTVKDLIGFGLQVAKGMKYLASKKFVHRDLAARNCMLDEKFTVKVADFG  
 50 ARDMYDKEYYSVHNKTGAKLPVKWMALESQTQKFTTKSDVWVSFGVVLWELMTR  
 GAPPYPDVNTFDITVLLQGRRLLOPEYCPDPLVEVMLK CWHPKAEAMRPSFSELVSRI  
 SAIFSTFGEHYHVHNATYVNVKCVAPYPSLLSSEDNADDEVDRPASFWETS

## SEQ ID NO. 12:

atggcgagtgtaattgttcgctacgatgcaagagtacacactcctcatttggatagcgcttgaagtgcccgaagtgtaagcccaactaca  
 gaaatgggttcaaatgaatctgtagactacgagctactttccagaagatcatttcaatctatcagaagggttcatgccgacaagt  
 gcagtagtccctcgacagacatgccccatcctaacatgtggggactctgatataccaggtccattactgcaaaatactgtccacattgacc  
 tcagtgcictaataccagagactggctccaggcagtcgacgatgtgattggcccgtagtgcattcaagaagtcacatag  
 gaagggggcatttgggtgtgtatatacgggactttgtggacaatgatggcaagaaattcactgtgctgtgaaatccttgaacagaatc  
 actgacataggagaagtttcccaatttggccggagggaatcatcatgaagatttttagtcatcccaatgtcctctcgtccttgggaatctg  
 cctgcgaagtgaagggtctccgcctgtgtgtcctaccatcatgaacaatggagacttcgaatttacttcgaaatgagactcataatcca  
 actgtaaaagacttatttggcttggcttcaatagtagccaaaggcatgaaatattctgcaagcaaaaagttgtccacagagacttggctgc  
 aagaatactgtatgctggatgaaatactcacagtcagggttgcatttggcttggccagagacatgtagataaagaatactatgtagtac  
 acaacaaaacaggttgcaagcgtccagtggaagtggaaggcttggaaagctctgcaaaactcaaaagttaccaccaagtcagatgtgtg  
 gtcccttggctgtcctctgggagctgatgacaaaggaggcccccacttactctgatgaaacacacttggatataactgtttacttggcga  
 agggagaagactctacacccgaatactgccagaccccttatgaagttaagctcaaaaatgctggcacccctaaagccgaaatgccc  
 ccactccttttgaactgtgtgccggatatacgaactctctctactttcattggggagcactatgtccatgtgaacgctactatgtgaacg  
 taaaatgtgtcgtccatatacctctgtgtgtcagaagataacgctgatgatgagggtgacacagaccagcctctctgtggagac  
 atca

[0046] As used herein, the abbreviation 'HGFRkd' refers to the catalytic domain of the HGFR, said domain beginning between residues 1051 and 1078 and terminating between residues 1341 and 1348 of the full-length protein [SEQ ID NO: 2]. According to certain embodiments of the present invention: (1) a methionine residue is added to the very N-terminal of the HGFRkd sequence [SEQ ID NO: 3]; (2) residues 1051-1349 of the hepatocyte growth factor receptor precursor wherein a methionine residue has been added to the very N-terminus of the sequence, the glycine at residue 1191 has been replaced by alanine, and the valine at position 1272 has been replaced by leucine [SEQ ID NO. 4]; (3) the valine at position 1272 is replaced by leucine [SEQ ID NO. 5]; (4) the methionine at residue 1250 is replaced by threonine [SEQ ID NO. 9]; a naturally occurring variant in hepatocellular carcinoma (HPRC); (5) the histidine at residue 1094 is replaced by arginine [SEQ ID NO. 15]; and/or (6) the tyrosine at residue 1230 is replaced by cysteine [SEQ ID NO. 16].

## SEQ ID NO. 3:

VHIDLALN PELVQAVQHV VIGPSSLIVH FNEVIGRGHF GCVYHGTLTD  
 NDGKKIHCAV KSLNRITDIG EVSQFLTEGI IMKDFSHPNV LSLGLICLRS EGSPLVVLPLY  
 MKHGDRLNFI RNETHNPTVK DLIGFGLQVA KGMKYLASKK FVHRDLAARN  
 CMLDEKFTVK VADFGRLARM YDKEYYSVHN KTGAKLPVKW MALESLOTQK  
 FTTKSDVWSF GVLLWELMTR GAPPYPDVNT FDTIVYLLQG RLLQPEYCP  
 DPLYEVMCLK WHPKAEMRPS FSELVSRISA IFSTFGEH

## SEQ ID NO.: 4

MVHIDLALN PELVQAVQHV VIGPSSLIVH FNEVIGRGHF GCVYHGTLTD  
 NDGKKIHCAV KSLNRITDIG EVSQFLTEGI IMKDFSHPNV LSLGLICLRS  
 EGSPLVVLPLY MKHGDRLNFI RNETHNPTVK DLIGFGLQVA KAMKYLASKK  
 FVHRDLAARN CMLDEKFTVK VADFGRLARM YDKEYYSVHN KTGAKLPVKW  
 MALESLOTQK FTTKSDVWSF GVLLWELMTR GAPPYPDVNT FDTIVYLLQG  
 RRLQPEYCP DPLYEVMCLK WHPKAEMRPS FSELVSRISA IFSTFGEH

SEQ ID NO.: 5

MVHIDLSALN PELVQAVQHV VIGPSSLIVH FNEVIGRGHF GCVYHGTLDD  
 NDGKKIHCAV KSLNRITDIG EVSQFLTEGI IMKDFSHPNV LSLLGICLRS  
 EGSPVLVLPY MKHGDRLNFI RNETHNPTVK DLIGFGLQVA KAMKYLASKK  
 FVHRDLAARN CMLDEKFTVK VADFGGLARM YDKEYYSVHN KTGAKLPVKW  
 MALESLOTQK FTTKSDVWSF GVLLWELMTR GAPPYPDVNT FDTIVYLLQG  
 RRLQPEYCP DPLYEVMLKC WHPKAEMRPS FSELVSISA IFSTFGEH

SEQ ID NO.: 9

MVHIDLSALN PELVQAVQHV VIGPSSLIVH FNEVIGRGHF GCVYHGTLDD  
 NDGKKIHCAV KSLNRITDIG EVSQFLTEGI IMKDFSHPNV LSLLGICLRS  
 EGSPVLVLPY MKHGDRLNFI RNETHNPTVK DLIGFGLQVA KAMKYLASKK  
 FVHRDLAARN CMLDEKFTVK VADFGGLARM YDKEYYSVHN KTGAKLPVKW  
 MALESLOTQK FTTKSDVWSF GVLLWELMTR GAPPYPDVNT FDTIVYLLQG  
 RRLQPEYCP DPLYEVMLKC WHPKAEMRPS FSELVSISA IFSTFGEH

SEQ ID NO.: 15:

MVHIDLSALN PELVQAVQHV VIGPSSLIVH FNEVIGRGHF GCVYHGTLDD  
 NDGKKIHCAV KSLNRITDIG EVSQFLTEGI IMKDFSHPNV LSLLGICLRS  
 EGSPVLVLPY MKHGDRLNFI RNETHNPTVK DLIGFGLQVA KAMKYLASKK  
 FVHRDLAARN CMLDEKFTVK VADFGGLARM YDKEYYSVHN KTGAKLPVKW  
 MALESLOTQK FTTKSDVWSF GVLLWELMTR GAPPYPDVNT FDTIVYLLQG  
 RRLQPEYCP DPLYEVMLKC WHPKAEMRPS FSELVSISA IFSTFGEH

SEQ ID NO.: 16

MVHIDLSALN PELVQAVQHV VIGPSSLIVH FNEVIGRGHF GCVYHGTLDD  
 NDGKKIHCAV KSLNRITDIG EVSQFLTEGI IMKDFSHPNV LSLLGICLRS  
 EGSPVLVLPY MKHGDRLNFI RNETHNPTVK DLIGFGLQVA KAMKYLASKK  
 FVHRDLAARN CMLDEKFTVK VADFGGLARM YDKEYYSVHN KTGAKLPVKW  
 MALESLOTQK FTTKSDVWSF GVLLWELMTR GAPPYPDVNT FDTIVYLLQG  
 RRLQPEYCP DPLYEVMLKC WHPKAEMRPS FSELVSISA IFSTFGEH

[0047] As used herein, the abbreviation 'pHGFR' refers to phosphorylated HGFR.

#### A. Peptides, Proteins and Antibodies

[0048] The present invention provides isolated peptide and protein molecules that are comprised of, consist of, or consist essentially of the amino acid sequences of the kinase peptides encoded by the nucleic acid sequence disclosed in the SEQ ID NO: 1, as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

[0049] As used herein, the terms "kinase", "kinase peptide" and "protein kinase" refer to enzymes that catalyze the transfer of a phosphate residue from a nucleoside triphosphate to an amino acid side chain in selected targets. The covalent phosphorylation in turn regulates the activity of the target protein. In addition, phosphorylation frequently acts as the signal that triggers a particular process or reaction, playing an integral part in cellular regulation and control mechanisms. Inappropriate or unregulated phosphorylation can result in errors in cell signaling and the associated cell cycle and regulation processes. Most protein kinases are highly substrate specific. Those that have the ability to phos-

phorylate numerous substrates frequently turn out to be oncogenes, genes that are associated with neoplastic transformation of a cell.

**[0050]** As used herein, the term "catalytically active form" refers to any form of peptides or proteins exhibiting intrinsic enzymatic activity. Preferably, the term "catalytically active form" refers to peptides or proteins capable of autophosphorylation.

**[0051]** As used herein, a peptide is said to be "isolated" or "purified" when it is free or substantially free of cellular material and/or free or substantially free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based primarily on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components.

**[0052]** In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), preferably less than about 20% other proteins, more preferably less than about 10% other proteins, or even more preferably less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

**[0053]** The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the kinase peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, preferably less than about 20% chemical precursors or other chemicals, more preferably less than about 10% chemical precursors or other chemicals, or even more preferably less than about 5% chemical precursors or other chemicals.

**[0054]** The isolated kinase described herein can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombination), or synthesized using known protein synthesis methods. For example, a nucleic acid molecule encoding the protein kinase is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

**[0055]** As mentioned above, the present invention also provides variants of the amino acid sequence of the peptides of the present invention, such as naturally occurring mature forms of the peptides, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can be generated using techniques that are known by those skilled in the fields of recombinant nucleic acid technology and protein biochemistry. In addition, such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

**[0056]** To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid 'identity' is equivalent to amino acid or nucleic acid 'homology'). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least about 40%, preferably about 50%, more preferably about 60%, even more preferably about 70%, still more preferably about 80%, and yet more preferably about 90% or more of the length of the reference sequence.

**[0057]** The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. See, e.g., *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York (1993); *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York (1991). In one embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into commercially available computer programs, such as GAP in the GCG software package, using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity

between two nucleotide sequences is determined using the commercially available computer programs including the GAP program in the GCG software package (Devereux, J., et al., *Nucleic Acids Res.* 12(1):387 (1984)), the NWS gap DNA CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into commercially available computer programs, such as ALIGN (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

**[0058]** The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using commercially available search engines, such as the NBLAST and XBLAST programs (version 2.0) of Altschul et al., *J. Mol. Biol.* 215:403-10 (1990). BLAST nucleotide searches can be performed, for example, with the NBLAST program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed, for example, with the XBLAST program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See, e.g., <http://www.ncbi.nlm.nih.gov>.

**[0059]** As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are preferably at least about 70-75%, more preferably at least about 80-85%, and even more preferably at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a peptide encoding nucleic acid molecule under stringent conditions as more fully described below. Peptides can readily be identified as having a high degree of (i.e., significant) sequence homology/identity to the peptides of the present invention. Full-length clones comprising one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the kinases of the present invention as well as being encoded by the same genetic locus as the kinase provided herein. Allelic variants of a peptide can readily be identified as having a high degree of sequence homology/identity to at least a portion of the peptide as well as being encoded by the same genetic locus as the kinase peptide provided herein.

**[0060]** Paralogues of a hepatocyte growth factor receptor kinase can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the HGFR, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are preferably at least about 60% or greater, and more preferably at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a HGFR encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

**[0061]** Orthologs of a kinase peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the kinase peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a kinase peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

**[0062]** Non-naturally occurring variants of the kinases of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the kinase. For example, one class of substitutions are conserved amino acid substitutions. Such substitutions are those that substitute a given amino acid in a kinase peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie et al., *Science* 247:1306-1310 (1990).

**[0063]** Variant kinases can be fully functional or may have reduced or decreased activity when compared to the wild-type protein. Fully functional variants may contain only conservative variations or variations in non-critical residues or in non-critical regions. Functional variants can also contain substitutions of similar amino acids, which result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncations or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

**[0064]** Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham et al., *Science* 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity, for example by measuring enzymatic activity. Sites that are critical for binding can also be

determined by structural analysis such as X-ray crystallography, nuclear magnetic resonance or photoaffinity labeling (Smith et al., *J. Mol. Biol.* 224:899-904 (1992); de Vos et al., *Science* 255:306-312 (1992)). Accordingly, the peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code; in which a substituent group is included; in which the polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol); or in which the additional amino acids are fused to the polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

**[0065]** The present invention further provides for functional, active fragments of the kinase peptides, in addition to proteins and peptides that comprise and consist of such fragments. As used herein, a fragment comprises at least about 8 or more contiguous amino acid residues from the protein kinase. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the kinase or could be chosen for the ability to perform a function, e.g. act as an immunogen. Preferred are fragments that are catalytically active and that have improved crystallographic properties as compared to the full-length, wild-type kinase. Such fragments will typically comprise a domain or motif of the kinase, e.g., active site. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs known and readily available to those of skill in the art (e.g., by PROSITE analysis- Hofmann et al., *Nucleic Acids Res.* 27:215-219 (1999); Bucher et al., *Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology AAAI Press, Menlo Park, 53-61* (1994)). For example, the fragment can comprise the HGFR intracellular domain [SEQ ID NO. 13].

#### SEQ ID NO. 13:

MGSELVRYDARVHTPHLDRLVSARSVSPITTEMVSNESVDYRATFPEDQFPNNSQNGS  
CRQVQVPLTDMSPILTSQDSISSPLLQNTVHIDLALNPVLQVQVHVIGPSSLIIVHF  
NEVIGRGHFGVCVYHGTLLDNDGKKIKHCAVKSLNRITDIGEVSQFLAEGIMKDFSHPN  
VLSLLGLCLRSEGLVLPYMKHGDLRNFIRNETHNPTVKDLIGFLQVAKGMKLYL

ASKKFVHRDLAARNCMLDEKFTVKVADFGLARDMYDKEYYSVHNKTAGKLPVKW  
MALESLOTQKFTTKSDVVSFGVLLWELMTRGAPPYDPVNTFDITVYLLQGRRLQPE  
YCPDPLYEVMLKCWHPKAEMRPSFSELVSRIASIFSTFIGEYHVHNATYVNVKCAV  
PPPSLLSSEDNADDEVDTRPASFWTES

**[0066]** A fragment is a variant peptide having an amino acid sequence that is entirely the same as part, but not all, of any amino acid sequence of any peptide of the invention. Fragments may be free standing or comprised within a larger peptide.

**[0067]** Polypeptides of the present invention also optionally contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature. Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, phenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter et al., *Meth. Enzymol.* 182: 626-646 (1990) and Rattan et al., *Ann. N.Y. Acad. Sci.* 663:48-62 (1992).

[0068] As used herein, "polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. The terms "peptide," "polypeptide," and "protein" are used interchangeably herein.

[0069] The peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a peptide operatively linked to a heterologous protein. "Operatively linked" indicates that the peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the kinase peptide. The two peptides linked in a fusion peptide are typically derived from two independent sources. Therefore, a fusion peptide comprises two linked peptides not normally found linked in nature. The two peptides may be from the same or different genome. In some uses, the fusion protein does not affect the activity of the peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions (i.e., Hi-tagged), MYC-tagged, and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant kinase peptides. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

[0070] A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments, which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see, Ausubel et al., *Current Protocols in Molecular Biology*, (1992)). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein, His-tag, or green fluorescent protein). A kinase peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the kinase peptide.

[0071] Herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one antibody combining site or binding domain, said binding domain or combining site formed from the folding of variable domains of an antibody molecule to form three dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigen epitope. The term encompasses immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, such as molecules that contain an antibody combining site or paratope. Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and portions of an immunoglobulin molecule, including those known in the art as Fab, FabB, F(abB)<sub>2</sub> and F(v).

## B. Nucleic Acids and Polynucleotides

[0072] The present invention provides isolated nucleic acid molecules that encode the functional or active kinases of the present invention. Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the kinase peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

[0073] The terms "nucleic acid molecule" and "polynucleotide" are used interchangeable in this application. These terms generally refer to any polyribonucleotide or polydeoxynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. These terms are intended to include DNA molecules (e.g., cDNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. These terms include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" and "nucleic acid molecule" as used herein refer to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the terms "polynucleotide(s)" and "nucleic acid molecule" also include DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The terms "polynucleotide(s)" and "nucleic acid molecules" as employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic



of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

**[0074]** As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA or cDNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein, such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, is preferably substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatograph such as HPLC.

**[0075]** For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

**[0076]** The preferred classes of nucleic acid molecules that are comprised of the nucleotide sequences of the present invention are the full-length cDNA molecules and genes and genomic clones since some of the nucleic acid molecules provided herein are fragments of the complete gene that exists in nature. A description of how various types of these nucleic acid molecules can be readily made/isolated is provided herein.

**[0077]** Full-length genes or portions thereof may be cloned from known sequences using any one of a number of methods known in the art. For example, a method which employs XL-PCR (Perkin-Elmer, Foster City, Calif.) to amplify long pieces of DNA may be used. Other methods for obtaining full-length sequences are known in the art.

**[0078]** The isolated nucleic acid molecules can encode the active protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to an active form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature active protein by cellular enzymes.

**[0079]** Once a full-length gene is cloned, portions of the gene can be obtained using techniques known to those of ordinary skill in the art. The isolated nucleic acid molecules include, but are not limited to, the sequence encoding the active kinase alone or in combination with coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the active kinase, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

**[0080]** Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA, obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

**[0081]** The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention and that encode obvious variants of the kinase proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

**[0082]** According to certain embodiments of the present invention, mutations to HGFR are utilized. For example, the tyrosine at residue 12130 is replaced with cysteine [SEQ ID NO. 10; a germline mutation in HPRC]; the methionine 1250 at residue 1250 is replaced with threonine [SEQ ID NO. 11]; and/or the histidine at residue 1094 is replaced with

arginine [SEQ ID NO. 14].

# SEQ ID NO. 10:

atggtccaca ttgacctcag tgcctctaaat ccagagctgg tccaggcagt gcagcatgta  
gtgatgggc ccagtgcct gattgtgcat tcaatgaag tcatagggaag agggcatttt  
gggttggtat atcatgggac ttgttggac aatgatggca agaaaattca ctgtgctgtg  
aaatccctga acagaatcac tgacatagga gaagtgtccc aattctgac cgagggaatc  
atcatgaaag attttatgca tcccaatgic ctctgctcc tgggaatctg cctgcgaagt  
gaagggtctc cgctgggtgt cctaccatac atgaacaatg gagatcttcg aaatttcatt  
cgaaatgaga ctcataatcc aactgtaaaa gatcttattg gctttgtct tcaagtggcc  
aaaggcatga aatacttgc aagcaaaaaa ttgtccaca gagacttggc tgcagaagaac  
tgtatgctgg atgaaaaatt cacagtcacg gttgctgatt ttggtctgc cagagacatg  
tgtataaag aatactatag tttacacac aaacagggtg caaagctgcc agtgaagtgg  
atggcttgg aaagtctgca aactcaaaag ttaccacca agtcagatgt gtgttccttt  
ggcgtctcc tctggagct gatgacaaga ggagccccc cttatctga cgtaaacacc  
tttgataaa ctgtttactt gttgcaaggg agaagactcc tacaaccga atactccca  
gacctctat atgaagtaat gctaaaatgc tggcacccta aagccgaat gcgccatcc  
ttttgaac tgggtcccc gatatacgc atcttcta ctttattgg ggagcac

# SEQ ID NO 11:

atggtccaca ttgacctcag tgcctctaaat ccagagctgg tccaggcagt gcagcatgta  
gtgatgggc ccagtgcct gattgtgcat tcaatgaag tcatagggaag agggcatttt  
gggttggtat atcatgggac ttgttggac aatgatggca agaaaattca ctgtgctgtg  
aaatccctga acagaatcac tgacatagga gaagtgtccc aattctgac cgagggaatc  
atcatgaaag attttatgca tcccaatgic ctctgctcc tgggaatctg cctgcgaagt  
gaagggtctc cgctgggtgt cctaccatac atgaacaatg gagatcttcg aaatttcatt  
cgaaatgaga ctcataatcc aactgtaaaa gatcttattg gctttgtct tcaagtggcc  
aaaggcatga aatacttgc gagcaaaaaa ttgtccaca gagacttggc tgcagaagaac  
tgtatgctgg atgaaaaatt cacagtcacg gttgctgatt ttggtctgc cagagacatg  
tgtataaag aatactatag tttacacac aaacagggtg caaagctgcc agtgaagtgg  
accgcttgg aaagtctgca aactcaaaag ttaccacca agtcagatgt gtgttccttt  
ggcgtctcc tctggagct gatgacaaga ggagccccc cttatctga tgaacacc  
tttgataaa ctgtttactt gttgcaaggg agaagactcc tacaaccga atactccca  
gacctctat atgaagtaat gctaaaatgc tggcacccta aagccgaat gcgccatcc  
ttttgaac tgggtcccc gatatacgc atcttcta ctttattgg ggagcac

## SEQ ID NO 14:

atggctcaca ttgacctcag tgctctaaat ccagagctgg tccaggcagt gcagcatgta  
 gtgattgggc ccagtagcct gatttgcat tcaatgaag tcataggaag agggcatttt  
 ggttgatgat atcgtgggac ttgttgac aatgatgca agaaaattca ctgtctgtg  
 aaatcctga acagaatcac tgacatagga gaagttccc aattctgac cgagggaatc  
 atcatgaag attttagtca tccaatgct ctctcgctc tgggaatcgt cctcgaaat  
 gaagggtctc cgttggtggt cctaccatac atgaacatg gagatcttcg aaatttcatt  
 cgaatgaga ctcataatcc aactgtaaaa gatcttatg gctttgctc tcaagtagcc  
 aaaggatga aatatcttg aagcaaaaaa ttgtccaca gagacttggc tgcagaatac  
 tgtatgctgg atgaaaaat cacagtcaag gttgctgatt ttggtcttcg cagagacatg  
 tatgataaag aatactatag tgtacacac aaacagggtg caaagctgcc agtgaagttg  
 atgctcttgg aaagtctgca aactcaaaag ttaccacca agtcagatgt gtggtccttt  
 ggcgtcgtcc tcgtggagct gatgacaaga ggagccccc cttatctga cgtaaacacc  
 ttgtatataa ctgttactt gttcgaagg agaagactcc tacaacccga atactgccca  
 gacccttat atgaagtaat gctaaaatgc tggcacctca aagccgaat gcgccatcc  
 ttcttgaac tgggtcccc gatatcagc atctctcta ctttcttgg ggagcac

[0083] A fragment comprises a contiguous nucleotide sequence greater than about 8 or more nucleotides. Further, a fragment could be at least about 30, preferably at least about 40, more preferably at least about 50, and even more preferably at least about 100 or more nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

[0084] A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or more consecutive nucleotides.

[0085] Orthologs, homologs, and allelic variants can be identified using methods known in the art. As described above, these variants comprise a nucleotide sequence encoding a peptide that is preferably about 60-65%, more preferably about 70-75%, and even more preferably at least about 90-95% or more homologous to the nucleotide sequence provided in SEQ ID NO: 1 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions to the nucleotide sequence shown in SEQ ID NO: 1 or a fragment of the sequence.

[0086] As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least about 50%, and more preferably at least about 55% or more, homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 65%, preferably at least about 70%, and more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989), which is hereby incorporated by reference in its entirety. One example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C.

[0087] As used herein, the term "hybridizes under moderate conditions" is intended to describe conditions for hybridization and washing which are less severe than those described above for stringent conditions. Such moderate conditions are known to those skilled in the art and can be found in *Molecular Cloning, A laboratory manual*, J. Sambrook, E.F. Fritsch, T. Maniatis, Cold Spring Harbor Press Book 2 Chapter 9. One example of moderate conditions is hybridization in 6X SSC at room temperature, followed by 2x SSC and 0.1% SDS at 37°C.

[0088] The nucleic acid molecules of the present invention are useful for probes, primers, and chemical intermediates, and in biological assays. For example, the nucleic acid molecules can be used as hybridization probes for cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described herein and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides described herein. The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in SEQ ID NO: 1. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' non-

coding regions. However, as discussed, fragments are not to be construed as those which may encompass fragments disclosed prior to the present invention. Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a kinase protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated.

[0089] The nucleic acid molecules of the present invention are useful for producing peptides for use in crystallization studies, drug discovery, and drug design. The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize anti-sense molecules of desired length and sequence. The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations. In addition, the nucleic acid molecules are useful for expressing antigenic portions of the proteins; for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods; for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein; for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides; for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides; and for making vectors that express part, or all, of the peptides.

[0090] Further, the nucleic acid molecules are useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in kinase protein expression relative to normal results.

[0091] *In vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

### C. Vectors and Host Cells

[0092] The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, that can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC. Various expression vectors can be used to express polynucleotides encoding an hHGFR kinase, such as (but not limited to) pET and pProEX. A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates. The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

[0093] Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

[0094] The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage  $\lambda$ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats. In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers. The term "operably linked" as used herein indicates that a gene and a regulatory sequence(s), such as a promoter, are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins or proteins which include transcriptional activation domains) are bound to the regulatory sequence(s).

[0095] In addition to containing sites for transcription initiation and control, expression vectors can also contain se-

quences necessary for transcription termination and, in the transcribed region, a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2001), which is hereby incorporated by reference in its entirety.

[0096] A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g., cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook et al., *supra*.

[0097] The regulatory sequence may provide constitutive expression in one or more host cells (i.e., tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are known to those of ordinary skill in the art.

[0098] The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are known to those of ordinary skill in the art.

[0099] The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using techniques known to those of ordinary skill in the art. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

[0100] As described herein, it may be desirable to express a peptide of the present invention as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of such peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith et al., *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., *Gene* 69:301-315 (1988)) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

[0101] Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California 119-128 (1990)). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli* (Wada et al., *Nucleic Acids Res.* 20:2111-2118 (1992)).

[0102] The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast, e.g., *S. cerevisiae*, include pYepSec1 (Baldari, et al., *EMBO J.* 6:229-234 (1987)), pMFA (Kurjan et al., *Cell* 30:933-943 (1982)), pJRY68 (Schultz et al., *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA). The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow et al., *Virology* 170:31-39 (1989)). In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B., *Nature* 329:840 (1987)) and pMT2PC (Kaufman et al., *EMBO J.* 6:187-195 (1987)). Each of the foregoing references is hereby incorporated by reference in its entirety.

[0103] The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. Preferred vectors include pET28a (Novagen, Madison, WI), pAcSG2 (Pharmingen, San Diego, CA), and pFastBac (Life Technologies, Gaithersburg, MD). The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found, for example, in Sambrook et al., *supra*.

[0104] The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of anti-

sense RNA. Thus, an anti-sense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this anti-sense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

[0105] The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells. Preferred host cells of the instant invention include *E. coli* and Sf9.

[0106] The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook et al., *supra*.

[0107] Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

[0108] In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

[0109] Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

[0110] While the active protein kinases can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

[0111] Where secretion of the peptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

[0112] It is also understood that depending upon the host cell used for the recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

[0113] The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a kinase protein or peptide that can be further purified to produce desired amounts of kinase protein or fragments. Thus, host cells containing expression vectors are useful for peptide production. Host cells are also useful for conducting cell-based assays involving the kinase protein or kinase protein fragments. Thus, a recombinant host cell expressing a kinase polypeptide of the invention is useful for assaying compounds that stimulate or inhibit kinase protein function. Host cells are also useful for identifying kinase protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant kinase protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native kinase protein.

#### D. Crystallization and Computer Methods for Model Building and Drug Design

[0114] Crystals of the polypeptides of the invention or ligand complexes of such polypeptides can be grown by a number of techniques including batch crystallization, vapor diffusion (either by sitting drop or hanging drop) and by microdialysis. Seeding of the crystals in some instances is required to obtain X-ray quality crystals. Standard micro and/or macro seeding of crystals may therefore be used. The HGFR-Compound 1 complex can be prepared as described below in reference to Example 2.

[0115] Once a crystal of the present invention is grown, X-ray diffraction data can be collected. X-ray diffraction data collection can be obtained using, for example, a MAR imaging plate detector. Crystals can be characterized by using X-rays produced in a conventional source (such as a sealed tube or a rotating anode) or using a synchrotron source.

[0116] Data processing and reduction can be carried out using programs such as HKL, DENZO, and SCALEPACK (Otwinowski and Minor, 1997, *Meth. Enzymol.* 276:307-326 (1997)). In addition, X-PLOR (Bruger, X-PLOR v.3.1 Manual, New Haven: Yale University, (1993)) or Heavy (T. Terwilliger, Los Alamos National Laboratory) may be utilized for bulk solvent correction and B-factor scaling. Electron density maps can be calculated using SHARP (La Fortelle, E. D. and

Bricogne G., *Meth. Enzymol.* 276:472-494 (1997)) and SOLOMON. Molecular models can be built into this map using O (Jones, T. et al., *ACTA Crystallogr.* A47:110-119 (1991)), XTALVIEW (Scripps Research) or QUANTA96 (Accelrys, Inc. San Diego). Refinement can be done using XPLOr (Brunger, "X-PLOR-A System for X-ray Crystallography and NMR," Yale University Press, New Haven, Conn), using the free R-value to monitor the course of refinement.

[0117] Once the three-dimensional structure of a crystal comprising HGFR or an HGFR-complex is determined, a potential ligand (antagonist or agonist) is examined through the use of computer modeling using a docking program such as FelixDock (Tripos, St. Louis, MO), GRAM (Medical Univ. Of South Carolina), DOCK3.5 and 4.0 (Univ. Calif. San Francisco), Glide (Schrödinger, Portland, OR), Gold (Cambridge Crystallographic Data Centre, UK), FLEX-X (BioSolveIT GmbH, Germany); AGDOCK (in-house software from Agouron Pharmaceuticals; Gehlhaar et al., *Chemistry & Biol.* 2:317-324), Hex (Ritchie, D and Kemp, G., *Proteins: Struct. Funct. & Genet.* 39:178-194), or AUTODOCK (Scripps Research Institute). This procedure can include computer fitting of potential ligands to the HGFR substrate-binding domain to ascertain how well the shape and the chemical structure of the potential ligand will complement or interfere with the HGFR substrate-binding domain (Bugg et al., *Scientific American* Dec.92-98 (1993); West et al., *TIPS* 16:67-74 (1995)). Computer programs can also be employed to estimate the attraction, repulsion, and steric hindrance of the ligand to the HGFR binding domain. Generally the tighter the fit (e.g., the lower the steric hindrance, and/or the greater the attractive forces) the more potent the potential drug will be since these properties are consistent with a tighter-binding constant.

[0118] "Binding domain" also referred to as "binding site", "binding pocket", "substrate-binding site," "catalytic domain," or "substrate-binding domain," refers to a region or regions of a molecule or molecular complex, that, as a result of its shape, can associate with another chemical entity or compound. Such regions are of significant utility in fields such as drug discovery. The association of natural ligands or substrates with binding pockets of their corresponding receptors or enzymes is the basis of many biological mechanisms of action. Similarly, many drugs exert their biological effects via an interaction with the binding pockets of a receptor or enzyme. Such interactions may occur with all or part of the binding pocket. An understanding of such interactions can lead to the design of drugs having more favorable and specific interactions with their target receptor or enzyme, and thus, improved biological effects. Therefore, information related to ligand binding with the HGFR substrate-binding site is valuable in designing potential modulators of HGFR. Further, the more specificity in the design of a potential drug the more likely that the drug will not interact with other similar proteins, thus, minimizing potential side effects due to unwanted cross interactions.

[0119] Initially, a potential ligand could be obtained by screening a random chemical library. A ligand selected in this manner could be then be systematically modified by computer-modelling programs until one or more promising potential ligands are identified. Such analysis has been shown to be effective in the development of HIV protease inhibitors (Larn et al., *Science* 263:380-384 (1994); Wlodawer et al., *Ann. Rev. Biochem.* 62:543-585 (1993); Appelt, *Perspectives in Drug Discovery and Design* 1:109-128 (1993); Erickson, *Perspectives in Drug Discovery and Design* 1:109-128 (1993)). Such computer modeling allows the selection of a finite number of rational chemical modifications, as opposed to the countless number of essentially random chemical modifications that could be made, and of which any one might lead to a useful drug. Each chemical modification requires additional chemical steps, which while being reasonable for the synthesis of a finite number of compounds, quickly becomes overwhelming if all possible modifications needed to be synthesized. Thus, through the use of the structure coordinates disclosed herein and computer modeling, a large number of these compounds can be rapidly screened on the computer monitor screen, and a few likely candidates can be determined without the laborious synthesis of untold numbers of compounds.

[0120] Once a potential ligand (agonist or antagonist) is identified it can be either selected from commercial libraries of compounds or alternatively the potential ligand may be synthesized *de novo*. As mentioned above, the *de novo* synthesis of one or even a relatively small group of specific compounds is reasonable in the art of drug design. The prospective drug can be tested in the binding assay exemplified below to test its ability to bind to the HGFR substrate-binding domain. The effect of the prospective drug on HGFR activity can also be determined using the assay described herein or other HGFR assays known in the art.

[0121] When a suitable compound is identified, a supplemental crystal can be grown which comprises a protein-ligand complex formed between the HGFR domain and the compound. Preferably the crystal effectively diffracts X-rays allowing the determination of the atomic coordinates of the protein-ligand complex to a resolution value of about 3.0 Å or less, more preferably about 2.0 Å or less. Molecular Replacement Analysis can be used to determine the three-dimensional structure of the supplemental crystal.

[0122] Molecular replacement involves using a known three-dimensional structure as a search model to determine the structure of an identical or closely related molecule or protein-ligand complex in a new crystal form. The measured X-ray diffraction properties of the new crystal are compared with those calculated from a search model structure to compute the position and orientation of the protein in the new crystal. Computer programs that can be used for this purpose include: X-PLOR, EPMR (Kissinger et al. *Acta Cryst.* D55:484-491 (1999)), ProLSQ and AMORE (J. Navaza, *Acta Crystallographica* ASO, 157-163 (1994)). Once the position and orientation are known an electron density map can be calculated using the search model to provide X-ray phases. Thereafter, the electron density is inspected for

structural differences and the search model is modified to conform to the new structure. Using this approach, it is possible to use the claimed structure to solve the three-dimensional structures of any such HGFR polypeptide-ligand complex. Other computer programs that can be used to solve the structures of such HGFR crystals include X-site, QUANTA, INSIGHT, ARP/wARP, and ICM.

[0123] For all of the drug design strategies described herein further refinements to the structure of the drug will generally be necessary and can be made by the successive iterations of any and/or all of the steps provided by the aforementioned strategies.

[0124] Another aspect of the invention involves using the structure coordinates generated from the HGFR-ligand complex to generate a three-dimensional shape. This is achieved through the use of commercially available software that is capable of generating three-dimensional graphical representations of molecules or portions thereof from a set of structure coordinates.

[0125] It will be readily apparent to those of skill in the art that the numbering of amino acids in other isoforms of HGFR may be different than that set forth for herein. Corresponding amino acids in other isoforms of HGFR are easily identified by inspection of the amino acid sequences, for example, through the use of commercially available homology software programs.

[0126] The amino acids of the HGFR domain of the polypeptides of the invention are described herein and are defined by a set of structure coordinates set forth in Table 1. The terms "structure coordinates" or "atomic coordinates" refer to Cartesian coordinates derived from mathematical equations related to the patterns obtained on diffraction of a monochromatic beam of X-rays by the atoms (scattering centers) of a protein or protein-ligand complex in crystal form. The diffraction data are used to calculate an electron density map of the repeating unit of the crystal. The electron density maps are then used to establish the positions of the individual atoms of the enzyme or enzyme complex.

Table 1

ATOM	1	CB	LEU	1062	-5.656	32.472	9.323	1.00	42.24		C
ANISOU	1	CB	LEU	1062	4988	5484	5576	-719	-875	457	C
ATOM	2	CG	LEU	1062	-4.343	31.747	9.028	1.00	45.82		C
ANISOU	2	CG	LEU	1062	6417	5073	5918	773	-2605	-523	C
ATOM	3	CD1	LEU	1062	-3.291	31.922	10.117	1.00	49.80		C
ANISOU	3	CD1	LEU	1062	6543	6638	5741	1783	-2327	-3085	C
ATOM	4	CD2	LEU	1062	-4.608	30.252	8.793	1.00	61.21		C
ANISOU	4	CD2	LEU	1062	8940	5485	8834	2	-3713	-1422	C
ATOM	5	C	LEU	1062	-4.927	34.835	9.088	1.00	32.29		C
ANISOU	5	C	LEU	1062	2270	5351	4646	198	-130	1309	C
ATOM	6	O	LEU	1062	-3.746	35.184	9.314	1.00	35.90		O
ANISOU	6	O	LEU	1062	3421	4477	5741	-1230	-770	879	O
ATOM	7	N	LEU	1062	-6.957	34.298	10.421	1.00	35.32		N
ANISOU	7	N	LEU	1062	2492	5438	5488	363	-111	3338	N
ATOM	8	CA	LEU	1062	-5.610	33.834	10.010	1.00	31.18		C
ANISOU	8	CA	LEU	1062	1981	4568	5319	-249	-1111	1419	C
ATOM	9	N	VAL	1063	-5.635	35.320	8.052	1.00	33.88		N
ANISOU	9	N	VAL	1063	4004	4197	4671	563	-652	829	N
ATOM	10	CA	VAL	1063	-4.971	36.369	7.245	1.00	37.92		C
ANISOU	10	CA	VAL	1063	4464	4071	5871	117	-1511	1514	C
ATOM	11	CB	VAL	1063	-5.771	36.763	6.002	1.00	40.52		C
ANISOU	11	CB	VAL	1063	5622	4212	5561	183	-1748	1365	C
ATOM	12	CG1	VAL	1063	-5.271	38.081	5.394	1.00	45.84		C
ANISOU	12	CG1	VAL	1063	5460	5756	6200	-558	-1740	2571	C
ATOM	13	CG2	VAL	1063	-5.720	35.689	4.922	1.00	49.72		C
ANISOU	13	CG2	VAL	1063	5692	6405	6795	826	-2035	-304	C
ATOM	14	C	VAL	1063	-4.710	37.538	8.200	1.00	39.48		C
ANISOU	14	C	VAL	1063	3451	5518	6030	-897	-1094	793	C
ATOM	15	O	VAL	1063	-3.699	38.229	8.177	1.00	40.62		O
ANISOU	15	O	VAL	1063	3403	5114	6917	-825	-1890	2272	O
ATOM	16	N	GLN	1064	-5.682	37.725	9.084	1.00	37.95		N